# How are research data governed at Japanese repositories? A knowledge commons perspective

Kai Nishikawa

Graduate School of Library, Information and Media Studies, University of Tsukuba, Tsukuba, Japan

# Abstract

**Purpose** – The purpose of this paper is to survey how research data are governed at repositories in Japan by deductively establishing a governance typology based on the concept of openness in the context of knowledge commons and empirically assessing the conformity of repositories to each type.

**Design/methodology/approach** – The fuzzy-set ideal type analysis (FSITA) was adopted. For data collection, a manual assessment was conducted with all Japanese research data repositories registered on re3data.org.

**Findings** – The typology constructed in this paper consists of three dimensions: openness to resources (here equal to research data), openness to a community and openness to infrastructure provision. This paper found that there is no case where all dimensions are open, and there are several cases where the resources are closed despite research data repositories being positioned as a basis for open science in Japanese science and technology policy.

**Originality/value** – This is likely the first construction of the typology and application of FSITA to the study of research data governance based on knowledge commons. The findings of this paper provide practitioners insight into how to govern research data at repositories. The typology serves as a first step for future research on knowledge commons, for example, as a criterion of case selection in conducting in-depth case studies.

Keywords Research data, Open science, Data governance, Fuzzy-set ideal type analysis, Knowledge commons, Research data repository

Paper type Research paper

# 1. Introduction

In Japan, research data repositories have been positioned as a basis for opening research data (Cabinet Office, 2019; Science Council of Japan, 2016), which is a core element of the open science movement, along with open access to research papers. Since the G8 Science Ministers Statement (G8 Science Ministers, 2013), Japan has promoted open science as a primary issue of science and technology policy (Cabinet Office, 2016). The Science Council of Japan, a representative organisation of Japanese scientists, made policy recommendations on the need for governing research data repositories "based on data strategies for open/closed data" (Science Council of Japan, 2016, p. iii). In 2019, the Japanese Cabinet Office (2019) also stressed the importance of considering an open-closed strategy when governing research data in its Integrated Innovation Strategy.

However, these documents seem to be built on insufficient empirical findings and do not provide a specific model of how to govern research data at repositories. Few empirical studies discuss Japanese research data repositories (Ikeuchi, 2019; Ikeuchi and Itsumura, 2016), and, to the best of our knowledge, there is no survey on how research data are governed at

© Kai Nishikawa. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at: http:// creativecommons.org/licences/by/4.0/legalcode.

Research data governed at Japanese repositories

# 837

Received 17 March 2020 Revised 12 June 2020 Accepted 17 June 2020



Aslib Journal of Information Management Vol. 72 No. 5, 2020 pp. 837-852 Emerald Publishing Limited 2050-3806 DOI 10.1108/AJIM-03-2020-0072 repositories in Japan although such knowledge would be essential for developing more specific strategies or policies.

As a critical work on research data, Borgman (2015) is helpful for approaching this gap, understanding research data as *knowledge commons* for analysing data practices. The term *commons* here refers to "a resource shared by a group of people that is subject to social dilemmas" (Hess and Ostrom, 2007, p. 3), and the term *knowledge commons* specifically pays attention to *knowledge*, which is defined as follows: "a broad set of intellectual and cultural resources (...) information, science, knowledge, creative works, data, and so on" (Frischmann *et al.*, 2014, p. 3).

Hardin's (1968) famous metaphor of "the tragedy of the commons" has led to the assumption that unrestricted knowledge eventually collapses because free riders, i.e. self-interested users who benefit from the commons without contributing to its maintenance, diminish the incentives of inventors and creators of that knowledge. The narrative has further evolved that there are only two solutions to avoid this tragedy: the privatisation of knowledge by setting intellectual property rights or the intervention of a state-based authority such as public subsidy (Madison *et al.*, 2018). However, knowledge commons studies have disproved these assumptions (e.g. Frischmann *et al.*, 2014; Hess and Ostrom, 2007; Madison *et al.*, 2018; Ostrom and Hess, 2007), and Borgman (2015) also states that research data as knowledge commons can be governed by a community consisting of diverse stakeholders not depending on property rights or a state-based authority if the community has suitable governance model (Borgman, 2015; Hess and Ostrom, 2007).

The purpose of this paper is to survey the governance of research data at repositories in Japan from a knowledge commons perspective. While Borgman (2015) used the early stage of knowledge commons studies as a framework for analysing research data, this paper draws on recent developments in this field and sheds new light on the governance of research data at repositories. To achieve this purpose, this paper deductively constructs a typology of the governance of research data at repositories in Japan and then empirically assesses those repositories' actual conformity to each type using fuzzy-set ideal type analysis (FSITA).

Although this paper only focuses on Japanese repositories, its methods will be applicable to repositories in other regions. This paper provides insight for practitioners involved in research data into how to govern research data at repositories. In addition, the methods and findings in this paper serve as a first step for future research on research data and knowledge commons studies.

Section 2 introduces this paper's theoretical foundations in detail and surveys relevant previous works. Section 3 briefly outlines FSITA, and in Section 4, the typology of the governance of research data at repositories is deductively constructed. Section 5 describes the method to measure the conformity of actual repositories to each ideal type, and Section 6 presents its results. Section 7 discusses the important findings and limitations of this paper. Finally, the conclusions, research and practical implications, and direction of future research are discussed in Section 8.

# 2. Theoretical foundation

#### 2.1 Knowledge commons

Early studies of knowledge commons have tended to focus on the characteristics of resources as economic goods, classifying types of goods into a two-by-two matrix based on exclusion potential and subtractability (Hess and Ostrom, 2003, 2007; Ostrom and Ostrom, 1977). Borgman (2015) also followed this tendency and constructed a typology of data included in a knowledge commons, based on the matrix of goods. Exclusion here refers to the degree of difficulty involved in excluding people from using a resource. Subtractability, also known as rivalry, is the degree to which someone's use of a resource limits its use by others. Although knowledge has been considered as a typical example of public goods, with low subtractability

AIIM

72.5

and for which exclusion is difficult, Borgman (2015) noted that research data can be categorised into any of the four categories, regarding data repositories as common-pool resources, with high subtractability and for which exclusion is difficult. Incidentally, Fecher *et al.* (2015) concluded that research data are not a knowledge commons since research data does not meet the definition of knowledge commons: "A good that can be accessed by everyone and whose consumption is non-rivalry" (Fecher *et al.*, 2015, p. 19). However, this paper does not accept this definition because their understanding of knowledge commons is as equal to public goods and knowledge commons is not limited to a single type of goods (Borgman, 2015; Frischmann *et al.*, 2014; Hess and Ostrom, 2007).

In the past decade, governance has become a central issue in current knowledge commons studies (Frischmann et al., 2014; Madison et al., 2018; Nishikawa, 2019), which redefined knowledge commons to focus on governance: "Knowledge commons is thus shorthand for the institutionalised community governance of the sharing and, in some cases, creation, of information, science, knowledge, data, and other types of intellectual and cultural resources" (Frischmann et al., 2014, p. 3). As earlier the fragmentary nature of knowledge commons studies limited their contributions to empirical research, these studies proposed a systematic research framework with standardised research questions to generate and integrate empirical findings on knowledge commons (Frischmann et al., 2014; Madison et al., 2010a, 2016, 2018). This approach ultimately aims to detect governance-related factors, such as the basic principles to govern shared resources sustainably or the differing degrees of governance models' effectiveness, through a comparative analysis. This new systematic approach avoids the narrower distinction of goods in a knowledge commons and widely captures diverse cases and enables collaboration among diverse disciplines (Madison et al., 2018). Unlike Borgman (2015), this paper follows the above definition and focuses on governance, not their characteristics as goods, and considers repositories as *infrastructure*, not as resources. By focusing on the governance of research data, it seems more natural to consider that the main resources for repositories are research data and repositories themselves are infrastructure—an element of the governance—as explained in Section 2.4.

# 2.2 Governance

Generally, the term governance refers to "all process of governing, whether undertaken by a government, market, or network, whether over a family, tribe, formal or informal organisation, or territory, and whether through laws, norms, power, or language" (Bevir, 2012, p. 1). In the context of knowledge commons, governance of knowledge commons can be explored from the three perspectives: (1) degrees of openness, (2) general governance structures and (3) rules and norms for a particular action arena (Frischmann *et al.*, 2014). This paper adopts the degrees of openness perspective because it is the most suitable for this paper's cross-sectional research design and is a critical element of research data governance at repositories in the Japanese open science policy mentioned in Section 1.

Vassilakopoulou *et al.* (2016, 2019) also investigated the governance and openness of genetic data from a perspective of knowledge commons, but they viewed the openness of data governance based on the matrix of types of goods, mentioned in Section 2.1, emphasising the characteristics of data as goods. On the other hand, this paper assumes that openness relates not only to data but also other dimensions. The following section discusses openness in more detail.

### 2.3 Open/openness

In general, the term open/openness in relation to data is defined by the Open Knowledge Foundation (2015) as: "Open means anyone can freely access, use, modify, and share for any purpose"; they further clarified related terms and the detailed requirements that open work must satisfy and the exceptions.

Research data governed at Japanese repositories

AJIM 72,5

840

In knowledge commons studies, openness is divided into openness as applied to (1) resources and (2) a community (Frischmann *et al.*, 2014; Madison *et al.*, 2010a). Openness to resources is openness in relation to knowledge resources, thus being compatible with the Open Knowledge Foundation's (2015) definition. The term *community* here means a group of people who share and create knowledge resources. Openness to a community thus "describes an individual's capacity to relate to that community as a contributor, manager, or user of resources that make up the knowledge commons" (Frischmann *et al.*, 2014, p. 29).

Here, it is important to note that openness in the context of knowledge commons is not the ideal goal, unlike open movements such as FAIR Data Principles (Wilkinson *et al.*, 2016). Although there is a tendency to regard openness as a panacea to govern knowledge resources, current knowledge commons studies do not assume that pure open governance is superior to closed options, such as privatisation by intellectual property rights (Frischmann *et al.*, 2014). This paper agrees and does not assume that the more open, the better.

#### 2.4 Infrastructure

Despite infrastructure's traditional association with physical components such as roads, electrical grids and telecommunications networks, it now includes non-physical aspects, called intellectual or knowledge infrastructure (Borgman, 2015; Frischmann, 2013). The term infrastructure in the context of knowledge commons studies (Madison *et al.*, 2016; de Rosnay and Musiani, 2016) means a foundation that supports any interaction, such as the creation and sharing of resources, by a community and is sometimes used interchangeably with knowledge commons itself (Borgman, 2015). However, as mentioned in Section 2.1, this paper considers research data repositories as infrastructure, following Vassilakopoulou *et al.* (2019).

Furthermore, Morell (2010, 2014) indicated that openness to the provision of infrastructure affects openness to resources and a community. For example, a community can change policies, such as terms of use, if the participation in the infrastructure provision is open; however, infrastructure providers can freely change policy without a community's permission if the infrastructure provision is closed. Infrastructure providers here refers to those who "technically, legally, and economically sustain" infrastructure (Morell, 2014, p. 299), and in this paper, they correspond to a repository's responsible institutions or specific board members. As Morell (2014) stated, infrastructure has rarely been considered in previous works, but this paper regards it as a dimension of repositories' research data governance in addition to openness to resources and a community.

# 2.5 Typology

Typologies are critical in the process of sciences, for example, developing theories, measuring changes and functioning as a first step for causal inference within quantitative research (Collier *et al.*, 2012). A typology is a tool to obtain an overall picture of a phenomenon. According to Ebbinghaus (2012), typologies can be divided into (1) a real typology developed by inductive/empirical methods, such as cluster analysis and (2) an ideal typology developed by deductive/theoretical methods, such as the fuzzy-set analysis. An ideal typology is specifically useful for future empirical research, e.g. as a basis to produce hypotheses or as a case selection criterion (Ebbinghaus, 2012). An ideal typology lives up to Weber's *ideal type* as "formed by the one-sided accentuation of one or more points of view and by the synthesis of a great many diffuse, more or less present and occasionally absent concrete individual phenomena, which are arranged according to those one-sidedly emphasised viewpoints into a unified analytical construct" (Weber, 2017, p. 90).

Current knowledge commons studies assume comparative institutional analysis (CIA) as the overall research design (Frischmann *et al.*, 2014; Madison *et al.*, 2016), and CIA often uses typologies to map institutional variations (Lange and Meadwell, 1991). Pampel *et al.* (2013) provided a typology of research data repositories. However, their typology was based on the attributes of research data and responsible institutions and did not focus on knowledge commons studies.

As mentioned in Section 1, this paper aims to construct an ideal typology using FSITA. While previous works provided typologies of knowledge commons' governance (e.g. Benkler, 2013; de Rosnay and Musiani, 2016), this paper appears to be the first to construct a typology of the governance of research data at repositories from the perspective of openness in the context of knowledge commons.

# 3. Fuzzy-set ideal type analysis (FSITA)

In this paper, FSITA is used to develop a general picture of research data governance at repositories in Japan. FSITA originates from qualitative comparative analysis (QCA), a data analysis method based on set theory (Ragin, 2000; Rihoux and Ragin, 2009; Schneider and Wagemann, 2012). FSITA is an application of QCA to establish a typology deductively, and empirically measure the conformity of cases to *a priori* established types (Ciccia and Verloo, 2012; Kvist, 1999, 2007). Several studies have adopted FSITA to construct typologies or conduct comparative analyses of institutions, policies or regimes (An and Peng, 2016; Ciccia, 2017; Ciccia and Verloo, 2012; Hudson and Kuehner, 2013; Huh *et al.*, 2018; Kowalewska, 2017; Kvist, 1999, 2007).

FSITA generally comprises four steps (An and Peng, 2016; Ciccia and Verloo, 2012). First, theoretically relevant dimensions of the ideal types are defined. As mentioned in Section 2.5, ideal types here function as an analytical tool to measure how closely empirical cases fit these categories, which represent reality as a kind of model but do not necessarily exist in an empirical sense (Weber, 2017). The number of logically possible ideal types is  $2^k$ , where *k* is equal to the number of selected dimensions.

Second, these dimensions are expressed as fuzzy sets, which have a degree of membership. In this step, researchers must determine *empirical indicators* to measure the dimensions. Once empirical values are obtained through empirical indicators, they are translated into 0 to 1 fuzzy-set membership scores. This process, called *calibration* in QCA, requires: (1) the definition of full membership (1), (2) the definition of full non-membership (0) and (3) the definition of the most ambiguous point that distinguishes membership or non-membership (0.5). These definitions are called *qualitative anchors*, which determine a case's fuzzy-set membership.

Third, the fuzzy-set membership score of each ideal type is calculated by using two principles of fuzzy-set theory: *minimum principle* and *logical negation*. The *minimum principle* states that a case's conformity to an ideal type is the minimum value of the involved sets' membership scores. For example, if a case's membership scores of A, B and C sets are respectively 0.6, 0.8 and 0.1, the case's membership score of the ideal type A\*B\*C (where \* = logical and) is the minimum value, 0.1 (the score of C). The *logical negation*, also called the complement, is a "set that contains all those cases that are not members in the original set" (Schneider and Wagemann, 2012, p. 323). As far as a case is not fully in (or out of) a certain set, the case has partial membership of both the set and its *logical negation*. The membership score of the logical negation is 1 – the membership score of the original set. For example, if a case has a 0.8 membership score on set A, its membership score for set a (the logical negation of set A) is 0.2 (=1-0.8).

Finally, researchers examine which ideal type a case belongs to. After the calculation with the two principles mentioned above, each case will have membership (membership score >0.5 and  $\leq$  1) only in one ideal type, based on the ideal type with the highest membership score. Note that a case rarely has full membership (membership score 1) in an ideal type because ideal types do not necessarily exist in reality.

Research data governed at Japanese repositories

FSITA can express a concept as a combination of several dimensions (sets), rather than as mutually independent, divisible variables (Kvist, 2007) and is good at dealing with small and medium-sized cases in a systematised way (Rihoux and Ragin, 2009). This paper adopts FSITA because these characteristics are suitable for analysing complex concepts such as openness.

# 4. Defining ideal types

As mentioned in Section 2, in the context of knowledge commons, the governance of knowledge resources such as research data can be explored from the concept of openness, and openness can be viewed as applied to resources (R) and a community (C) (Frischmann et al., 2014; Madison et al., 2010). Furthermore, this paper adds openness to infrastructure provision (I) as another dimension since the level of I affects the level of openness to resources and a community (Morell, 2010, 2014).

Table 1 depicts 8 ( $=2^3$ ) logically possible ideal types that are constructed from the three dimensions and indicates the ideal typology of the governance of research data at repositories based on the perspective of openness.

# 5. Operationalising and calibrating the fuzzy sets

# 5.1 Data source

In this paper, the empirical data were collected through a manual assessment of Japanese research data repositories registered on the global registry re3data.org (https://www.re3data. org/), which crosses academic disciplines and countries.

This paper first collected 56 cases of repositories registered as Japanese by searching re3data.org using the "Browse by country" tab. These cases were then narrowed down to 37 cases according to the following exclusion criteria:

- (1) When the repository itself cannot be accessed due to broken links, etc., or when it is accessible, but its primary function is unavailable due to obsolete technology, etc.
- (2) When it is duplicated.
- (3) When the repository has no digitised research data.
- (4) When the repository is substantially operated by foreign organisations because Japanese organisations have only a subsidiary role. The decision is based on the information recorded in re3data.org that indicates institutions' responsibility in the repository ("Type (s) of responsibility") and the information provided on the repository's web page or in the relevant literature.
- (5) When the repository is a collection of links to individual databases and does not have governing bodies or policies.

	Ideal type	Openness to resources ( <i>R</i> )	Openness to community ( <i>C</i> )	Openness to infrastructure provision ( <i>I</i> )
Table 1.         Ideal typology of the governance of research data at repositories (or eight logically possible ideal types)	R*C*I r*C*I R*c*I R*C*i r*c*i r*C*i R*c*i r*c*i	$ \begin{array}{l} R \ (\text{open}) \\ r \ (\text{closed}) \\ R \ (\text{open}) \\ r \ (\text{closed}) \\ r \ (\text{closed}) \\ R \ (\text{open}) \\ r \ (\text{closed}) \\ r \ (\text{closed}) \\ \end{array} $	C (open) $C (open)$ $c (closed)$ $C (open)$ $c (closed)$ $C (open)$ $c (closed)$ $c (closed)$	I (open) I (open) I (open) i (closed) I (open) i (closed) i (closed) i (closed) i (closed)

AIIM

72.5

Manual assessment here refers to the following: accessing repositories, examining policies, registering as a user and downloading research data. In addition, the relevant literature for each repository was also obtained. The data were collected from October to December 2019.

5.2 Setting empirical indicators

This paper developed indices of each dimension to measure their empirical values as all dimensions are complex concepts and cannot be measured by a single indicator. Each dimension is measured from the perspective of an external user who is not a member of the repository's provision body. The scores were manually assigned by the author.

Table 2 illustrates the index of openness to resources (*R*) using four empirical indicators (*R* Ind.1–4), which follow the Open Definition version 2.1 (Open Knowledge Foundation, 2015) and are based on the following two previous indices of open data on government: the Global Open Data Index (Open Knowledge Foundation, 2017) and the Open Data Barometer Leaders Edition (World Wide Web Foundation, 2017). The indicators that are suitable for the context of research data were selected from these two indices. Here, *R* was measured from architectural/technological (*R* Ind.1–2), legal/normative (*R* Ind.3), and economic aspects (*R* Ind.4) and these aspects theoretically correspond to the four constraints—architecture, law, norm and market—proposed by Lessig (1998). Each indicator in *R* uses a graded scale scoring system that could appropriately classify and measure the condition of repositories, while the previous indices were scored using a binary system. The specific scoring method was determined by conducting a preliminary case study on the repositories selected in Section 5.1.

The *R* score is assigned according to the following procedures:

purpose, no charge (Open Knowledge Foundation, 2015)

- (1) When a repository has a standard usage policy, the score was assigned based on the policy.
- (2) When there is no standard policy, and a repository contains multiple datasets with different usage conditions, the main dataset was determined according to the repository's mission, goals and objectives, or background, and the score was assigned focusing on the main dataset.
- (3) When there is no standard policy and a single main dataset cannot be determined, the score was assigned focusing on the most open dataset in the repository.

In addition, on *R* Ind.3, whether it is "open license" or not was determined based on the open license requirements as proposed by the Open Knowledge Foundation (2015).

Table 3 illustrates the index of openness to a community (C). C is further divided into three aspects: a contributor, manager and user (Frischmann *et al.*, 2014; Madison *et al.*, 2010).

Indicator	Score
<i>R</i> Ind.1: Is the data in open and machine- readable file formats? <i>R</i> Ind.2: Is the data downloadable at once? <i>R</i> Ind.3: Is the data openly licensed*/in the public domain?	<ul> <li>0 - Neither open nor machine-readable, 1 - Not open but machine-readable, or vice versa, 2 - Open and machine-readable</li> <li>0 - Not downloadable, 1 - Downloadable but not in bulk,</li> <li>2 -Downloadable at once</li> <li>0 - Copyrighted/no policy (unknown), 1 - Not open licensed,</li> <li>2 - Open licensed/in public domain</li> </ul>
<i>R</i> Ind.4: Is the data free of charge?	0 - Charged, 1 - Not charged
<b>Note(s)</b> : * The term "open license" is redistribution, modification, separation.	a license that permits or meets the following conditions: use, compilation, non-discrimination, propagation, application to any

Table 2. The index of openness to resources (R)

Research data governed at Japanese repositories

AJIM	Indicator	Score	
72,0	<i>C</i> Ind.1: Possibility of uploading data	0 - Closed to non-members/no explicit rule, 1 - Moderated registration*, 2 - Automatic registration**, 3 - No registration	
	<i>C</i> Ind.2: Possibility of deciding the terms of use of data	0 - Impossible/no explicit rule, 1 - Possible	
844	<i>C</i> Ind.3: Registration or requirement to use data	0 - Closed to non-members, 1 - Moderated registration , 2 - Automatic registration, 3 - No registration	
<b>Table 3.</b> The index of opennessto a community (C)	<b>Note(s)</b> : * The term "moderated registration" here refers to a registration system where a moderator can filter those who register to become part of the community ** The term "automatic registration" here refers to a registration system that does not require any filter to become part of the community (Morell, 2010)		

In the context of this paper, a contributor refers to a researcher who provides research data to a repository and is measured by CInd.1. As the manager aspect involves different levels, it is partly measured by C Ind.2 and partly by I Ind.2. C Ind.3 measures the user aspect. The scoring method was determined based on the preliminary case study as with R.

On C Ind.1 and 3. a distinction was established between automatic registration (score 2) and moderated registration (score 1), because automatic registration is "non-discriminatory" (Frischmann, 2013) or "symmetric" (Benkler, 2013, 2014), which does not differentiate based on a register's identity or purpose of use; moderated registration, however, may impose discriminatory or "asymmetric" (Benkler, 2013, 2014) restrictions. This distinction is also based on the indicator of openness to participation in the platform proposed by Morell (2010). On C Ind.3, when there are multiple datasets requiring different conditions regarding registration, the score is assigned according to the same procedures with R. In addition, if registration is required for a specific use, such as commercial use, or if registration is not required to access the data but is required to download the data, registration was determined to exist. On the other hand, if registration is not mandatory, it was determined that there was no registration.

Table 4 illustrates the index of openness to infrastructure provision (*I*). Both the indicators and each scoring method were adopted from the qualitative values of two axes concerning infrastructure provision proposed by Morell (2010), with a slight change in the description of the score illustrated in Table 4 to fit the context of this paper. Although Morell (2010) focused not on research data but online creation communities (OCCs), her method can be applied to this paper because OCCs are a broader concept that can involve research data repositories, as demonstrated by her case concerning scientific knowledge, such as the Public Library of

	Indicator	Score
	<i>I</i> Ind.1: Level of freedom/autonomy of participants from the infrastructure provider	0 - Proprietary software and copyright license/unknown, 1 - Use of FLOSS* but not copyleft license**, 2 - Use of FLOSS and copyleft license
	<i>I</i> Ind.2: Possibility of participation in provision body	<ul> <li>0 - By becoming a member of a commercial (for-profit) body,</li> <li>1 - By becoming a member of a non-profit body,</li> <li>2 - By fulfilling certain criteria or meeting requirements,</li> <li>3 - Participation by self-selection (everybody who wants to join)</li> </ul>
<b>Table 4.</b> The index of opennessto infrastructureprovision ( <i>l</i> )	Note(s): * The term "FLOSS" is an abbreviation of free software and open-source software ** The term "copyleft license" refers to a license the license as the original	of "Free/Libre and Open Source Software", and refers to both at requires derivative works to be distributed under the same

Science. This dimension involves two aspects: the forkability of the infrastructure (*I* Ind.1) and the possibility for people to participate in the infrastructure provision (*I* Ind.2; Morell, 2010, 2014).

*I* Ind.1 measures whether the infrastructure can be forked or not. *Forking* is copying source code from an open-source software program and developing and distributing a new program independently. If the repository uses FLOSS software and copyleft licenses, the repository is forkable, and the contributor or user can easily move research data in it elsewhere or relaunch a similar repository, separately from the provider of the original repository. However, if the repository uses proprietary or copyrighted software, the community is locked into the repository. In other words, the infrastructure provision of the repository can be considered more open when forkable, and vice versa.

*I* Ind.2 measures the possibility of participating in the provision body. As mentioned in Section 2.4, participation in the provision body means involvement in decision making about the infrastructure. Therefore, the governance of the infrastructure provision was considered closed when it was difficult for an outsider to become a member of the infrastructure providers, and it is open when participating in the provision body is easy for everyone.

### 5.3 Setting qualitative anchors

Table 5 illustrates the qualitative anchors mentioned at Section 3. Regarding all of the indices in this paper, anchors 0 were set at the minimum value (0), anchors 1 were set at the maximum value and anchors 0.5 were set at the middle value of each index. Usually, the use of statistical parameters such as the mean or median as qualitative anchors is not encouraged since these values depend on the statistical distribution of cases and do not correspond to qualitative differences among them (Rihoux and Ragin, 2009; Schneider and Wagemann, 2012). Despite these recommendations, this paper used the middle value of each index as 0.5 anchors for the following reasons: (1) the value used in this paper does not function as the statistical parameter of empirical cases but only as the middle value of the indices and (2) both the maximum and minimum values of the indices respectively correspond to fully open and fully non-open states; therefore, adopting the middle values of the indices as qualitative anchors corresponding to the most ambiguous state (0.5) was justified.

To translate empirical data into the fuzzy-set membership score, this paper adopted the qualitative approach, which assigns prepared fuzzy scores to each dimension according to qualitative anchors, as the empirical data collected in this paper was quasi-interval-scale data and therefore, the semi-automatic methods of the calibrations seemed less appropriate (Schneider and Wagemann, 2012). More specifically, this paper used a four-value fuzzy-set schema (Rihoux and Ragin, 2009; see Table 6).

Dimension	Fully in (1.00 anchor)	Neither in nor out (0.50 anchor)	Fully out (0.00 anchor)	
R	7 (the maximum of <i>R</i> )	3.5 (the middle value of $R$ )	0 (the minimum of <i>R</i> )	Table 5.Dimensions andqualitative anchors
C	7 (the maximum of <i>C</i> )	3.5 (the middle value of $C$ )	0 (the minimum of <i>C</i> )	
I	5 (the maximum of <i>I</i> )	2.5 (the middle value of $I$ )	0 (the minimum of <i>I</i> )	

Fuzzy value	Verbal description			
1 0.67 0.33 0	Fully in More in than out More out than in Fully out	Four-value fuzzy-set schema: fuzzy-set membership scores and their verbal descriptions		

Research data governed at Japanese repositories

AJIM 72,5

# 6. Results

Table 7 illustrates the fuzzy-set membership scores of each repository for each ideal type—in other words, repositories' degree of conformity to each ideal type. A case is considered to belong to the ideal type where the membership score is written in italic. The number of the cases belonging to each ideal type is as follows: three cases to  $R^*c^*I$ , nine cases to  $R^*C^*i$ , one case to  $r^*C^*i$ , twenty-two cases to  $R^*c^*i$ , and two cases to  $r^*c^*i$ . No cases were assigned to  $R^*C^*I$ ,  $r^*C^*I$ , or  $r^*c^*I$ . The empirical value of each case for each indicator is illustrated in the Appendix.

# 846

	Case no	Case name	R*C*I							
	1	ASTER j-space systems	0.33	0.33	0.33	0.33	0.33	0.33	0.67	0.33
	2	Human genetic variation repository	0.33	0	0.33	0.33	0	0	0.67	0
	3	Intermagnet	0.33	0.33	0.33	0.33	0.33	0.33	0.67	0.33
	4	Tropical atmosphere ocean project	0.33	0	0.33	0.33	0	0	0.67	0
	5	International service of geomagnetic indices	0.33	0	0.33	0.33	0	0	0.67	0
	6	Brain transcriptome database	0.33	0	0.33	0.33	0	0	0.67	0
	7	Data centre for aurora in NIPR	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.67
	8	ADS	0.33	0.33	0.33	0.33	0.33	0.33	0.67	0.33
	9	jPOSTrepo	0.33	0.33	0.33	0.67	0.33	0.33	0.33	0.33
	10	National institute of polar research science database	0.33	0	0.33	0.33	0	0	0.67	0
	11	Life science database archive	0.33	0.33	0.33	0.67	0.33	0.33	0.33	0.33
	12	Spectral database for organic compounds	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.67
	13	DIAS	0.33	0.33	0.33	0.67	0.33	0.33	0.33	0.33
	14	Japanese genotype-phenotype archive	0.33	0.33	0.33	0.67	0.33	0.33	0.33	0.33
	15	International mouse phenotyping consortium	0.33	0	0.67	0.33	0	0	0.33	0
	16	Nobevama radio polarimeters	0.33	0.33	0.33	0.33	0.33	0.33	0.67	0.33
	17	SOAP	0.33	0.33	0.33	0.33	0.33	0.33	0.67	0.33
	18	World data centre for geomagnetism, Kyoto	0.33	0.33	0.33	0.33	0.33	0.33	0.67	0.33
	19	ASTER IPL	0.33	0	0.33	0.33	0	0	0.67	0
	20	PDBi	0.33	Õ	0.33	0.67	Õ	Õ	0.33	Õ
	21	DNA data bank of Japan	0.33	Õ	0.33	0.67	Õ	Õ	0.33	Õ
	22	World data centre for cosmic rays	0.33	0.33	0.33	0.33	0.33	0.33	0.67	0.33
	23	GlyTouCan	0.33	0.33	0.33	0.67	0.33	0.33	0.33	0.33
	24	SMOKA science archive	0.33	0.33	0.33	0.33	0.33	0.33	0.67	0.33
	25	WDC for ionosphere and space weather	0.33	0.33	0.33	0.33	0.33	0.33	0.67	0.33
	26	World data centre for greenhouse gases	0.33	0.33	0.33	0.67	0.33	0.33	0.33	0.33
	27	Informatics research data repository	0.33	0.33	0.33	0.33	0.33	0.33	0.67	0.33
	28	LIMIN CTR	0.33	0.33	0.33	0.33	0.33	0.33	0.67	0.33
	29	Pig expression data explorer	0.33	0	0.67	0.33	0.00	0.00	0.33	0.00
	30	CURATOR	0.33	0.33	0.33	0.33	0.33	0.67	0.33	0.33
	31	DARTS	0.33	0	0.33	0.33	0.00	0	0.67	0.00
	32	Autophagy database	0.33	õ	0.33	0.33	Õ	Õ	0.67	0
	33	Dartmouth flood observatory	0.00	033	0.00	0.00	033	033	0.67	033
Table 7.       Fuzzy membership	34	Kyoto encyclopedia of genes and genomes	0.33	0.33	0.33	0.33	0.33	0.33	0.67	0.33
scores of the	35	Plant organelles database 3	0.33	0.33	0.33	0.67	0.33	0.33	0.33	0.33
dote at repositories in	36	Ial TFR MetaCat service	0.33	0.33	0.33	0.33	0.33	0.33	0.67	0.33
ideal types	37	FANTOM	0.33	0.00	0.67	0.33	0.00	0.00	0.33	0.00

# 7. Discussion

The goal of this paper was to provide a general picture of how research data are governed at repositories in Japan. The results measure the conformity of repositories to each ideal type. Before interpreting the results, it is important to remember that this paper does not assume that a more open state is desirable for governing research data at repositories. In the context of both research data in Japan and knowledge commons, many studies have suggested that the preferred governance type may vary depending on the situation (Cabinet Office, 2019; Frischmann *et al.*, 2014; Madison *et al.*, 2010a, 2016; Nishikawa, 2019; Science Council of Japan, 2016).

Interestingly, this paper reveals that there are cases where not even resources are open, although research data repositories are positioned as a foundation for promoting open science in Japan (Cabinet Office, 2019; Science Council of Japan, 2016). This seems to be because the current requirements of openness to resources—especially with regard to architectural/technological aspects—were not understood when these repositories were first established, and they have not since been fundamentally updated. The results further show that the ideal type with the most cases is  $R^*c^*i$ . Although it is theoretically implied that openness to a community and infrastructure provision influence how repositories are used (Frischmann *et al.*, 2014; Madison *et al.*, 2010a; Morell, 2010, 2014), open science policy in Japan considers only openness to resources (Cabinet Office, 2019; Science Council of Japan, 2016). This paper suggests that in addition to policy documents, actual repositories consider only openness to resources.

The  $r*C^*i$  is the most characteristic ideal type among types to which actual cases belong. The  $r*C^*i$  refers to governance where only a community is open, but resources and infrastructure provision are closed. Only CURATOR (case no. 30: https://opac.ll.chiba-u.jp/da/ curator/?lang=1), which is the only institutional repository involved in this paper, fell into this ideal type. With CURATOR, even those who do not belong to the provision body (Chiba University) can provide their data to CURATOR if they can obtain permission from the director of the university library, making the C dimension open. However, many resources within CURATOR are under copyright protection because the same usage policy connected with the research paper applies to the data, and resources are generally not machine-readable and cannot be downloaded at once. Moreover, external users cannot participate in the infrastructure provision because CURATOR has adopted a proprietary software (Asoshina, 2005) and its provision body is restricted to the university.

If other Japanese institutional repositories were analysed, they also would likely belong to the  $r^*C^*i$  because CURATOR was the first institutional repository in Japan and appears to have been a model for subsequent repositories. Although some institutional repositories use open-source software instead of proprietary software, their scores for I are still expected to be low because many of them are operated exclusively by universities or research institutes, making it difficult for outsiders to participate in the provision body. Their scores for *R* will also be low because many Japanese organisations have not yet developed a data policy (Ikeuchi, 2019), and the same usage policy that applies to the research paper is thus likely to apply to data in repositories other than CURATOR. If an institutional repository were to be used as a basis for opening research data, it would be necessary to develop policies for data that are separate from those for research papers.

A form of knowledge commons remains far from formulation (Madison *et al.*, 2010b); Nishikawa (2019) stated that knowledge commons is an inclusive term for a specific type of governance of knowledge resources and it is assumed that there are variations of knowledge commons. To the best of our knowledge, the typology in this paper explicitly shows these variations for the first time and provides clues to their formulation. In this regard, the question now arises: Can all ideal types—even cases where resources are closed—be called knowledge commons? As it turns out, a case that includes r would also be knowledge Research data governed at Japanese repositories

commons, because knowledge commons is not the same as open access (Hess and Ostrom, 2007). All repositories in this paper at least allow external users to access data and therefore seem to meet the general definition of knowledge commons to some degree (see Section 2.1).

This paper has some limitations. First, the number of datasets to be analysed was narrowed down to one to assign the score for R and partly for C, as stated in Section 5.2, but many repositories actually have multiple datasets with different conditions. Although this simplification is essential to provide an typology, the detailed information that each case originally had was lost as a result.

Second, some empirical indicators and scoring methods were originally developed in this paper, and the empirical scores were assigned by the author. Therefore, the author's subjectivity may have entered the analysis, although this seems inevitable because of the exploratory nature of this paper. To address this, this paper has tried to ensure transparency by clarifying the analytical procedure in as much detail as possible.

Third, unlike statistical methods, FSITA does not assume a population behind cases and re3data.org appears to register only a portion of Japanese repositories. It is thus not possible to say that the results of this paper accurately reflect the overall trend of research data governance at repositories in Japan, and there is a possibility that some repositories belong to ideal types that no case belonged to in this paper. For example, a repository related to citizen science would belong to the ideal type  $R^*C^*I$  since such a repository would have to make its community and infrastructure provision more open because of the importance of the role of people who do not belong to a specific research institute.

# 8. Conclusion

This paper offers a new way to examine the governance of research data at repositories, updating the perspective of Borgman (2015) and suggests that Japanese research data repositories emphasise only the openness of resources, at times without even opening their data. In addition, the ideal typology constructed here also represents variations of knowledge commons.

This paper is the first step towards enhancing the research into both the governance of research data and knowledge commons. The method used in this paper will be applicable to repositories in regions other than Japan. In addition, because there appear to be no other studies using FSITA in the field of knowledge commons, this paper is the first to demonstrate the effectiveness of applying FSITA to knowledge commons research. The typology and the results of the analysis could serve as a case selection criterion and a tool for producing hypotheses when conducting a comparative analysis or an in-depth case study, which is a basic research design for current knowledge commons research.

This paper could help decision-makers consider how to govern research data at repositories by exploring their current state beyond the simple open/closed dichotomy. The concept of openness involves several dimensions and all of them would influence how repositories are used. The open-closed strategy stated in policy documents should take these multiple perspectives into account.

Future research should proceed in two directions. First, a comparative analysis between different regions or academic disciplines with the indices of this paper could clarify the characteristics of research data governance at repositories. Second, an in-depth case study on repositories involved in this paper may improve knowledge about the governance mechanisms of knowledge commons and lead to its formulation.

#### References

An, M.Y. and Peng, I. (2016), "Diverging paths? A comparative look at childcare policies in Japan, South Korea and Taiwan", *Social Policy and Administration*, Vol. 50 No. 5, pp. 540-558, doi: 10.1111/spol.12128.

848

AIIM

72.5

- Asoshina, H. (2005), "How do we work for developing real IR?: practical knowledge from the CURATOR project", *Journal of Information Processing and Management*, Vol. 48 No. 8, pp. 496-508, doi: 10.1241/johokanri.48.496, [published in Japanese].
- Benkler, Y. (2013), "Commons and growth: the essential role of open commons in market economies", University of Chicago Law Review, Vol. 80 No. 3, pp. 1499-1555.
- Benkler, Y. (2014), "Between Spanish Huertas and the Open Road: A Tale of Two Commons?", in Madison, M.J., Strandburg, K.J. and Frischmann, B.M. (Eds), *Governing Knowledge Commons*, Oxford University Press, New York, pp. 69-98.
- Bevir, M. (2012), Governance: A Very Short Introduction, Oxford University Press, Oxford.
- Borgman, C.L. (2015), Big Data, Little Data, No Data: Scholarship in the Networked World, MIT Press, Cambridge, MA.
- Cabinet Office (2016), "The 5th Science and Technology Basic Plan", [published in Japanese], available at: https://www8.cao.go.jp/cstp/kihonkeikaku/5honbun.pdf (accessed 6 March 2020).
- Cabinet Office (2019), "Integrated innovation strategy 2019", [published in Japanese], available at: https://www8.cao.go.jp/cstp/togo2019\_honbun.pdf (accessed 6 March 2020).
- Ciccia, R. and Verloo, M. (2012), "Parental leave regulations and the persistence of the male breadwinner model: using fuzzy-set ideal type analysis to assess gender equality in an enlarged Europe", *Journal of European Social Policy*, Vol. 22 No. 5, pp. 507-528, doi: 10.1177/ 0958928712456576.
- Ciccia, R. (2017), "A two-step approach for the analysis of hybrids in comparative social policy analysis: a nuanced typology of childcare between policies and regimes", *Quality and Quantity*, Vol. 51 No. 6, pp. 2761-2780, doi: 10.1007/s11135-016-0423-1.
- Collier, D., LaPorte, J. and Seawright, J. (2012), "Putting typologies to work: concept formation, measurement, and analytic rigor", *Political Research Quarterly*, Vol. 65 No. 1, pp. 217-232, doi: 10.1177/1065912912437162.
- de Rosnay, M.D. and Musiani, F. (2016), "Towards a (de) centralisation-based typology of peer production", *TripleC*, Vol. 14 No. 1, pp. 189-207, doi: 10.31269/triplec.v14i1.728.
- Ebbinghaus, B. (2012), "Comparing welfare state regimes: are typologies an ideal or realistic strategy", *Proceedings of the European Social Policy Analysis Network*, in Edinburgh, UK, September 6-8, 2012, ESPAnet conference.
- Fecher, B., Friesike, S. and Hebing, M. (2015), "What drives academic data sharing?", *PloS One*, Vol. 10 No. 2, doi: 10.1371/journal.pone.0118053.
- Frischmann, B.M., Madison, M.J. and Strandburg, K.J. (2014), "Governing knowledge commons", in Madison, M.J., Strandburg, K.J. and Frischmann, B.M. (Eds), *Governing Knowledge Commons*, Oxford University Press, New York, pp. 1-43.
- Frischmann, B.M. (2013), Infrastructure: The Social Value of Shared Resources, Oxford University Press, New York.
- Hardin, G. (1968), "The tragedy of the commons", *Science*, Vol. 162 No. 3859, pp. 1243-1248, doi: 10.1126/science.162.3859.1243.
- Hess, C. and Ostrom, E. (2003), "Ideas, artifacts, and facilities: information as a common-pool resource", *Law and Contemporary Problems*, Vol. 66 Nos 1/2, pp. 111-145.
- Hess, C. and Ostrom, E. (2007), "Introduction: an overview of the knowledge commons", in Hess, C. and Ostrom, E. (Eds), Understanding Knowledge as a Commons: From Theory to Practice, MIT Press, Cambridge, MA, pp. 3-26.
- Hudson, J. and Kuehner, S. (2013), "Beyond indices: the potential of fuzzy set ideal type analysis for cross-national analysis of policy outcomes", *Policy and Society*, Vol. 32 No. 4, pp. 303-317, doi: 10.1016/j.polsoc.2013.10.003.
- Huh, T., Kim, Y. and Kim, J.H. (2018), "Towards a green state: a comparative study on OECD countries through fuzzy-set analysis", *Sustainability*, Vol. 10 No. 9, doi: 10.3390/su10093181.

Research data governed at Japanese repositories

AJIM 72,5	Ikeuchi, U. and Itsumura, H. (2016), "Data sharing policies in scholarly journals across different disciplines: a comparative analysis", <i>Journal of Japan Society of Library and Information</i> <i>Science</i> , Vol. 62 No. 1, pp. 20-37, doi: 10.20651/jslis.62.1_20, [published in Japanese].
	Ikeuchi, U. (2019), "Open data practices and perceptions of researchers in Japan", University of Tsukuba, Tsukuba, PhD thesis[published in Japanese]. doi: 10.15068/00158168.
850	Kowalewska, H. (2017), "Beyond the 'train-first'/'work-first' dichotomy: how welfare states help or hinder maternal employment", <i>Journal of European Social Policy</i> , Vol. 27 No. 1, pp. 3-24, doi: 10.1177/0958928716673316.
	Kvist, J. (1999), "Welfare reform in the Nordic countries in the 1990s: using fuzzy-set theory to assess conformity to ideal types", <i>Journal of European Social Policy</i> , Vol. 9 No. 3, pp. 231-252, doi: 10.1177/095892879900900303.
	Kvist, J. (2007), "Fuzzy set ideal type analysis", <i>Journal of Business Research</i> , Vol. 60 No. 5, pp. 474-481, doi: 10.1016/j.jbusres.2007.01.005.
	Lange, P. and Meadwell, H. (1991), "Typologies of democratic systems: from political inputs to political economy", in Wiarda, H.J. (Ed.), <i>New Directions in Comparative Politics</i> , Westview Press, Boulder, CO, pp. 82-117.

- Lessig, L. (1998), "The new Chicago school", The Journal of Legal Studies, Vol. 27 No. S2, pp. 661-691, doi: 10.1086/468039.
- Madison, M.J., Frischmann, B.M. and Strandburg, K.J. (2010a), "Constructing commons in the cultural environment", *Cornell Law Review*, Vol. 95 No. 4, pp. 657-709, doi: 10.31219/osf.io/76g93.
- Madison, M.J., Frischmann, B.M. and Strandburg, K.J. (2010b), "Reply: the complexity of commons", *Cornell Law Review*, Vol. 95 No. 793, pp. 839-850, doi: 10.31219/osf.io/c4x3h.
- Madison, M.J., Strandburg, K.J. and Frischmann, B.M. (2016), "Knowledge commons", Working Paper [No. 2016-28], University of Pittsburgh, Pittsburgh, PA, 23 September. available at: https:// papers.ssrn.com/sol3/papers.cfm?abstract\_id=2841456 (accessed 6 March 2020).
- Madison, M.J., Strandburg, K.J. and Frischmann, B.M. (2018), "Knowledge commons (2019)", Working Paper [No. 2018-39], University of Pittsburgh, Pittsburgh, PA, 12 December, available at: https://papers.ssrn.com/sol3/papers.cfm?abstract\_id=3300348 (accessed 6 March 2020).
- Morell, M.F. (2010), "Governance of online creation communities: provision of infrastructure for the building of digital commons", PhD thesis, European University Institute, Fiesole.
- Morell, M.F. (2014), "Governance of online creation communities for the building of digital commons: viewed through the Framework of Institutional Analysis and Development", in Madison, M.J., Strandburg, K.J. and Frischmann, B.M. (Eds), *Governing Knowledge Commons*, Oxford University Press, New York, pp. 281-312.
- Nishikawa, K. (2019), "A theoretical consideration on the systematisation of the study of the knowledge commons", *Japan Society of Information and Knowledge*, Vol. 29 No. 3, pp. 213-233, doi: 10.2964/jsik\_2019\_037, [published in Japanese].
- Open Knowledge Foundation (2015), "Open definition 2.1", available at: https://opendefinition.org/od/ 2.1/en/ (accessed 6 March 2020).
- Open Knowledge Foundation (2017), *Global Open Data Index Methodology*, available at: https://index. okfn.org/methodology/ (accessed 6 March 2020).
- Ostrom, E. and Hess, C. (2007), "A framework for analysing the knowledge commons", in Hess, C. and Ostrom, E. (Eds), Understanding Knowledge as a Commons: From Theory to Practice, MIT Press, Cambridge, MA, pp. 41-81.
- Ostrom, V. and Ostrom, E. (1977), "Public goods and public choices", in Savas, E.S. (Ed.), Alternatives for Delivering Public Services: Toward Improved Performance, Westview Press, Boulder, CO, pp. 7-49.

Pampel, H., Vierkant, P., Scholze, F., Bertelmann, R., Kindling, M., Klump, J., Goebelbecker, H.J., Gundlach, J., Schirmbacher, P. and Dierolf, U. (2013), "Making research data repositories visible: the re3data. org registry", *PloS One*, Vol. 8, p. 11, doi: 10.1371/journal.pone.0078080.

Ragin, C.C. (2000), Fuzzy-set Social Science, University of Chicago Press, Chicago, US-IL.

- Rihoux, B. and Ragin, C.C. (2009), Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques, Sage, Thousand Oaks, CA.
- Schneider, C.Q. and Wagemann, C. (2012), Set-theoretic Methods for the Social Sciences: A Guide to Qualitative Comparative Analysis, Cambridge University Press, Cambridge.
- Science Council of Japan (2016), Recommendations Concerning an Approach to Open Science that Will Contribute to Open Innovation, available at: http://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-23t230-en.pdf (accessed 6 March 2020).
- G8 Science Ministers (2013), "G8 science Ministers statements", available at: https://www.gov.uk/ government/news/g8-science-ministers-statement (accessed 6 March 2020).
- Vassilakopoulou, P., Skorve, E. and Aanestad, M. (2016), "Acommons perspective on genetic data governance: the case of BRCA data", *European Conference on Information System (ECIS) 2016 Proceedings in Istanbul, Turkey, 2016*, Association for Information System (AIS), Research Papers, 136.
- Vassilakopoulou, P., Skorve, E. and Aanestad, M. (2019), "Enabling openness of valuable information resources: curbing data subtractability and exclusion", *Information Systems Journal*, Vol. 29 No. 4, pp. 768-786, doi: 10.1111/isj.12191.
- Vis, B. (2007), "States of welfare or states of workfare? Welfare state restructuring in 16 capitalist democracies, 1985–2002", *Policy and Politics*, Vol. 35 No. 1, pp. 105-122, doi: 10.1332/ 030557307779657720.
- Weber, M. (2017), "Objectivity' in social science and social policy", in Shils, E.A. and Finch, H.A. (Eds), *The Methodology of the Social Sciences*, Routledge, London, pp. 49-112.
- Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.W., Santos, L.B.D., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., Hoen, P.A.C. 't, Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., Schaik, R. van, Sansone, S.A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., Lei, J. van der, Mulligen, E. van, Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B. (2016), "Comment: the FAIR guiding principles for scientific data management and stewardship", *Scientific Data*, Vol. 3, p. 160018, doi: 10.1038/sdata.2016.18.
- World Wide Web Foundation (2017), Open Data Barometer Leaders Edition Research Handbook, available at: http://opendatabarometer.org/doc/leadersEdition/ODB-leadersEdition-ResearchHandbook.pdf (accessed 6 March 2020).

#### **Corresponding author**

Kai Nishikawa can be contacted at: kai.nishikawa192000@gmail.com

Research data governed at Japanese repositories

	Append	lix
	ITotal	
	C Total	3332221253 424532123453413433303321413
	RTotal	くらすららくしてくややらのらののししくらららっかのての よののでしょう
	/Ind.2	
	/Ind.1	××××××××××××××××××××××××××××××××××××××
	C Ind.3	~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
	C Ind.2	00000001 00110000000000111010000000
	C Ind.1	000000000 00000000000000000000000000000
	$_{\rm Ind.4}^R$	
	R Ind.3	88-888888888888888888888888888888888888
	RInd.2	
	R Ind.1	~~~~~~~~~
	Case name	ASTER jspacesystems Human Genetic Variation Repository Intermagnet Tropical Atmosphere Ocean Project International Service of Geomagnetic Indices Brain Transcriptome database Data Centre for Aurora in NIPR ADS Program Institute of Polar Research Science database Life Science database Archive Spectral database for Organic Compounds DIAS Japanese Genotypephenotype Archive International Mouse Phenotyping Consortium Nobeyama Radio Polarimeters SolAP International Mouse Phenotyping Consortium Nobeyama Radio Polarimeters SolAP World Data Centre for Geomagnetism, Kyoto ASTER JPL World Data Centre for Geomagnetism, Kyoto ASTER JPL DNA Data Bank of Japan World Data Centre for Cosmic Rays GlyTouCan World Data Centre for Cosmic Rays GlyTouCan World Data Centre for Greenhouse Gases Informatics Research Data Repository UMIN CTR WORD for ionosphere and Space Weather World Data Centre for Greenhouse Gases Informatics Research Data Repository Pig Expression Data Explorer CURATOR DARTS Autophagy database Datrophage database JalTTER MetaCat Service FANTOM
1. irical values	Case no	33833333333333333333333333333333333333

# AJIM 72,5

852

Table A Raw emp